

Autoengaño e interacciones sociales

Self-Deception and Social Interactions

Carlos G. Patarroyo G. y Carlos A. Muñoz

(Universidad del Rosario y Universidad Agustiniana, Colombia)

carlos.munoz@uniagustiniana.edu.co

RESUMEN

En este artículo presentamos el modelo tradicional (que llamaremos canónico) que se ha usado para explicar el autoengaño y algunos de los problemas que ha de enfrentar. Presentaremos luego el que ha sido conocido como el modelo deflacionista o no intencionalista y defendemos que es una alternativa mejor para entender este fenómeno que el modelo canónico. Finalmente, presentaremos una limitación o problema que ha de enfrentar el modelo deflacionista y que, al ser tratada, obliga a hacer ciertas consideraciones y a mirarlo bajo una nueva luz que, lejos de ponerlo en duda o debilitarlo, termina por fortalecerlo. Esta limitación se basa en que el modelo ha sido pensado teniendo en mente sólo evidencia perceptual, pero no interacciones sociales. Defendemos que al incluir las interacciones sociales se hace necesario modificar levemente el modelo para dar cuenta de ellas y del comportamiento que los agentes a quienes adscribimos autoengaño tienen en estas relaciones. En cada una de las secciones de este texto acudiremos a ejemplos de la literatura que sirven para presentar con claridad las características de cada caso.

Palabras clave: autoengaño; modelo deflacionista; verdad; engaño; creencias contradictorias; relaciones sociales

ABSTRACT

In this paper we present the traditional (which we call canonical) model used to explain self-deception and show some of the problems it has to face. Then we present a different model known as the deflationary model or the non-intentional model and defend it is a better alternative to explain self-deception than the former. Finally, we present a problem the deflationary model has to face and that, when treated, leads to new considerations about the model and to see it under a new light. This problem, however, does not debilitate the model but makes it stronger in the end. We defend that the problem arises from the fact that the model has been thought on the light of perceptual evidence but not including social interactions. The inclusion of such interactions makes it necessary to modify slightly the model in order to accommodate them and the behavior of the self-deceived agents. In each of the sections of this paper we resort to literary examples that depict the essential characteristics of each case.

Keywords: Self-deception; deflationary model; truth; deception; contradictory beliefs; social interactions.

1. La caracterización canónica del autoengaño y sus paradojas

En su magistral escrito *Novela de ajedrez* Stefan Zweig presenta la historia de un prisionero, a quien identifica sencillamente como *el señor B.*, quien es encerrado en confinamiento solitario por largo tiempo. Sin nadie con quien hablar, sin nada que leer, sin ninguna distracción; está cerca de perder la cordura cuando se presenta la oportunidad robar un pequeño libro del bolsillo del abrigo de uno de los guardas que entra a su celda. Cuando el guarda abandona la celda, el prisionero ve, con algo de decepción, que se trata de un libro que reunía ciento cincuenta partidas de ajedrez de los maestros más famosos. No teniendo más distracción disponible, el prisionero se da a la tarea de imaginar las partidas en su mente. Con el tiempo logra reconstruir las partidas en su imaginación, ver las piezas y los movimientos en su mente. Llega hasta el punto de lograr memorizar las partidas a la perfección y ya no requiere del libro para reconstruirlas. Pasa las horas, interminables y lánguidas, reconstruyendo en su cabeza los enfrentamientos, pero con el transcurrir de los meses incluso esta se convierte en una aburrida actividad que ya no trae ninguna distracción. Es entonces que decide dar un paso adicional: se propone ya no reconstruir partidas, sino crear unas nuevas. Sin piezas, sin tablero, todo en su cabeza. El problema radica en que no tiene un contrincante, así que se decide a jugar contra sí mismo. El mismo Zweig comenta lo descabellado de este plan: “En el ajedrez, al ser un puro juego del pensamiento desligado por completo del azar, es lógicamente un absurdo querer jugar contra uno mismo” (65). El absurdo se relaciona con la complejidad que encierra una buena partida de ajedrez. Un buen jugador ha de trazar una estrategia para vencer a su contrincante pero, a la vez, esta estrategia ha de estar oculta pues, de descubrirla, el contrincante podría desarticularla con facilidad. Así, uno de los jugadores tiene en su mente un plan que el otro ha de desconocer, lo cual, por supuesto, parece ser imposible cuando alguien juega contra sí mismo:

Si una misma persona juega con las blancas y con las negras, se produce entonces una situación incongruente, en donde un mismo cerebro ha de saber y al mismo tiempo no saber, ha de

ser capaz de olvidar completamente cuando juega con las negras, lo que quería y pretendía cuando jugaba con las blancas. Un doble pensamiento como éste presupone en realidad una escisión absoluta de la conciencia, una capacidad de enfocar y desenfocar el cerebro como si fuera un aparato mecánico; querer jugar contra uno mismo representa, en definitiva, una paradoja tan grande en ajedrez como querer saltar sobre la propia sombra (65-6).

Sin embargo, esto es justamente lo logra hacer el Señor B. en el escrito de Zweig:

A partir del momento, en cambio, en que intenté ponerme a jugar contra mí mismo, empecé simultáneamente a desafiarme. Cada uno de mis dos yoes, mi yo blanco y mi yo negro, querían emularse, y los dos, independientemente, fueron presa de la ambición, de la impaciencia por vencer, por ganar. [...] Cada uno de mis dos yoes se regocijaba cuando el otro cometía un error, y se enojaba al mismo tiempo por la propia ineptitud (69).

Hemos escogido este ejemplo literario porque representa justamente las dos cosas que deseamos resaltar en esta sección; por un lado, lo que se conoce como la interpretación canónica del autoengaño en filosofía y, por el otro, lo compleja, desconcertante y paradójica que esta interpretación parece ser. Comencemos entonces por el primero de estos elementos. Para la interpretación canónica, el autoengaño no encierra, al menos en lo que a su estructura se refiere, ninguna complejidad; no se trata sino de la misma estructura del engaño interpersonal, pero en donde ya no hay dos personas sino sólo una (véase, por ejemplo, Bermúdez 2000; Funkhauser y Barrett 2016). La estructura del engaño interpersonal supone que hay dos sujetos, A y B, en donde:

- A cree que P (una proposición) es falso.
- A intenta que B crea que P es verdadero.
- A es exitoso en hacer creer a B que P es verdadero.

Aparte de los dos sujetos, hay dos elementos esenciales en esta estructura: (i) dos creencias opuestas (la creencia que tiene A: “P es verdadero” y la creencia que tiene B: “P es falso”), y (ii) la *intención de engañar* que tiene A. Si no hubiera esta intención no se podría hablar de engaño, sino de error, o malentendido, por ejemplo.

Ahora bien, como decíamos anteriormente, para la interpretación canónica, el autoengaño no es sino la misma estructura, sólo que ya no hay un sujeto B, sino sólo uno A en ambos roles:

- A cree que P es falso.
- A intenta que A crea que P es verdadero.
- A es exitoso en hacer creer a A que P es verdadero.

Y, tal como ocurría en el engaño interpersonal, aquí también hay creencias opuestas y contradictorias, así como la intención de engañar. Pero es justamente aquí en donde los problemas comienzan y son precisamente aquellos que el mismo Zweig hacía notar en su texto: ¿cómo puede ser siquiera concebible que alguien se intente engañar a sí mismo? Retomando el ejemplo del ajedrez ¿cómo puede alguien urdir una estrategia, un plan para vencerse a sí mismo y, a la vez, cómo puede este plan ser efectivo, para lo cual era necesario que uno mismo desconociera el plan a fin de caer presa de la estrategia? Habría que conocer (como engañador) el plan y desconocerlo (como engañado) al mismo tiempo. Como bien lo ha dicho Hugo Strandberg en su reciente obra *Self-Knowledge and Self-Deception* “Se puede decir que el autoengaño consiste en una combinación de saber y no-saber. De alguna manera sé algo –en mi capacidad de engañador– y de alguna manera no sé eso mismo –en mi capacidad de engañado–” (26)¹.

De manera similar, ¿cómo puede alguien creer simultáneamente una cosa y su opuesto? Es bien sabido que los seres humanos somos, en muchas de nuestras actitudes y creencias, contradictorios. Pero estas contradicciones no son tan directas como las que requiere la versión canónica del autoengaño. Alguien puede defender, por ejemplo, los derechos de los animales y ser, sin embargo, carnívoro. Para muchos esto es una contradicción, pero no es una contradicción directa y evidente *prima facie* sino una contradicción que se da entre lo que se cree y algunas de las acciones que reflejan que la creencia no acompaña a las acciones que deberían ser guiadas por ella. Lejos de ello en la

¹ Algunos han intentado caracterizar al autoengaño y, en particular, a este saber y a la vez no-saber, como una suerte de *insinceridad* (por ejemplo, Stokke, *Lying & Insincerity* 175 ss.). No perseguiremos esta ruta argumentativa por considerar que lo que se hace aquí es denominar “autoengaño” a un fenómeno completamente diferente.

versión canónica del autoengaño la contradicción es directa: el marido cree que su esposa le es infiel y, a la vez, cree que le ha sido siempre fiel, por poner un ejemplo muy utilizado en los textos filosóficos sobre el autoengaño.

No es nuestro interés en este texto desarrollar a fondo los intentos de solución de estas paradojas porque tampoco lo es defender la versión o interpretación canónica del autoengaño. Nos limitaremos, entonces, a resumir brevemente en qué han consistido algunos de los intentos más conocidos al respecto.

Una primera manera de intentar evitar una de las paradojas está en afirmar que, si bien hay una contradicción entre creencias opuestas en el mismo sujeto, ésta es posible porque una de las creencias es consciente mientras que la otra es inconsciente (Audi 94). Aparentemente esto parece resolver el problema, pero en realidad abre más preguntas y paradojas de las que soluciona. Para Greenwald una posición así debe poder responder antes que nada qué beneficio puede tener para un sujeto creer algo de manera inconsciente (51), y para Strandberg la pregunta es aún más básica, hay que responder primero qué se está entendiendo por “creer” algo (45). La respuesta no es para nada fácil; para ver la dificultad partamos de la manera como algunos han intentado responder la pregunta acerca de la creencia. Es fácil, como menciona Laurence Bonjour pensar que creer una proposición es “encontrarse en un estado consciente en el que el sujeto considera y afirma explícitamente la proposición en cuestión” (*Epistemology* 25). Sin embargo, el mismo autor reconoce de inmediato que hay muchas cosas que decimos que creemos y que no estamos considerando en este mismo momento, ni a las que estamos asintiendo presentemente. Por ello ofrece una doble caracterización de la creencia: por un lado están las creencias que llama *ocurrentes*, es decir, aquellas en las que estamos en un estado de consciencia y las consideramos y afirmamos y, por el otro, están aquellas que él llama *disposicionales*, que son aquellas en las que “la persona no tiene la proposición explícitamente en la mente, pero está dispuesta a aceptarla o asentir a ella si se le hiciera la pregunta” (29). Esto es algo que muchos otros han hecho también, con sutiles diferencias y otros nombres, pero siempre preservando lo esencial, por ejemplo, Buckwalter, Rose y Turri (2013) trazan la diferencia casi en los mismos términos, pero

llaman al primer tipo de creencias “grueso” (*thick*) y al segundo “delgado” (*thin*).

Si bien esto puede servir para incluir en el concepto de “creencia” aquellas proposiciones que un sujeto no tiene en su mente consciente en este momento, no parece funcionar para explicar cómo es posible que éstas sean contradictorias y estén en el mismo sujeto. Es fácil pensar que la distinción antes explicada sirva para decir que creo que mi abuela fue una mujer ejemplar, aún si no estoy pensando en mi abuela en este momento. Decir que esa creencia es disposicional o delgada permite que me sea adscrita; pero hay un trecho grande entre esto y decir que creo, ocurrentemente o de manera gruesa, que mi abuela era una mujer ejemplar y que, a la vez, creo disposicionalmente o de manera delgada, que no lo era. En el primer caso, aun si no estoy pensando en este momento en mi abuela, si alguien me pregunta si era una mujer ejemplar, asentiré. Esto es justamente lo que significa tener una creencia disposicional o delegada. Pero en el segundo caso parece que yo estoy pensando en mi abuela y en que ella era una mujer ejemplar, y alguien llega a preguntarme si creo que era una mujer ejemplar y le respondo que no, que no lo era. Parece que la inserción de la distinción entre tipos de creencias sirve para solucionar algunos problemas relativos a la epistemología, pero no para solucionar la paradoja de la versión canónica del autoengaño de las creencias opuestas. Esto, por supuesto, dista de ser un argumento contundente en contra de la versión canónica del autoengaño; pero no pretende serlo. Como dijimos anteriormente, no es nuestro propósito atacar contundentemente la versión canónica. Como suele ser común en filosofía, el debate sigue abierto. Para el lector interesado en ello, una defensa reciente de cómo se puede mantener el modelo del autoengaño basado en una mezcla de creencias conscientes e inconscientes es la ofrecida por Funkhauser y Barrett (2016); a la vez, esta propuesta puede verse como atentando contra el modelo de “transparencia” que se supone nuestras creencias suelen tener (Byrne 2018). Nosotros dejaremos aquí esta discusión para resumir brevemente otro intento de solución.

Para evitar recurrir a la alternativa de las creencias inconscientes recién vista, algunos autores han optado por ofrecer un modelo compartimentalizado de la mente humana. El filósofo Donald Davidson recurrió a esta estrategia:

El punto es que la gente puede tener, y de hecho en ocasiones tiene, creencias cercanamente relacionadas pero opuestas. Para ello debemos aceptar que puede haber fronteras entre distintas partes de la mente; postulo tal frontera en algún lugar entre cualquier par de creencias obviamente contradictorias. Estas fronteras no son descubiertas por introspección, son ayudas conceptuales para la descripción coherente de irracionalidades genuinas (*Problems of Rationality* 210).

De esta manera dos creencias opuestas y contradictorias podrían coexistir en la mente, pero porque nunca se enfrentan la una a la otra ya que cada una de ellas está encerrada en un compartimento diferente, hay una frontera que no se puede cruzar y que impide que la contradicción se haga evidente para el sujeto. Cruzar el límite, para Davidson, eliminaría una de las dos creencias y, por ende, la contradicción misma:

[L]as creencias contradictorias no pertenecen al mismo territorio, borrar la línea entre ellas destruiría una de las creencias. No veo ninguna razón obvia para suponer que los territorios deben estar cerrados a la conciencia –sea lo que eso signifique–, pero es claro que el agente no puede revisar la totalidad sin borrar las fronteras (211).

Allende de cómo funcionaría una solución como esta, el problema que se presenta es que se hace de inmediato difícil hablar de *un sujeto* que se engaña a sí mismo. Esta segmentación parece venir al costo de la unidad del sujeto. Justamente eso es lo que postula Amélie Okseberg-Rorty cuando habla del autoengaño. Para ella éste sólo es posible si se concibe al yo de dos maneras diferentes: por un lado, como una unidad y, por el otro, como una diversidad de compartimentos o segmentos que pueden incluir creencias contradictorias que jamás se encuentran de frente las unas con las otras:

La aparente dificultad del autoengaño proviene de la *superposición*, en vez de la *subsunción o complementariedad* de las dos imágenes, de los dos yoes. Cuando una imagen es proyectada sobre la segunda, el yo es engañado por sí mismo cuando sus subsistemas [...] se resisten a la corrección e integración disponible a ellos (*The Deceptive Self* 25).

¿Cómo funciona este modelo? No es para nada claro. Sólo siendo muy generales es posible hablar de esta *unidad en medio*

de la diversidad, pero al entrar en detalles parece que los problemas afloran. Se supone que hay una evidencia frente al sujeto, ¿por qué ella no borra las fronteras de las que habla Davidson? ¿Cómo entender esta “resistencia” de los subsistemas de los que habla Okseberg-Rorty? ¿Hay intencionalidad allí? También es difícil ver en esta propuesta la intención de engañarse que requiere el modelo canónico. ¿Engaña el sujeto a un subsistema, un subsistema engaña al otro? Nuevamente, no es nuestro deseo refutar propuestas como estas, pero creemos que es patentemente visible que se trata de intentos de alta complejidad, enrevesados y exigentes, tanto teórica como empíricamente. Creemos que hay un modelo explicativo para los casos que se suelen denominar de autoengaño que, al prescindir de algunas de las características del modelo canónico, evita tener que buscar enrevesadas explicaciones para hacer frente a las paradojas de las que el canónico adolece. Lo presentaremos a continuación.

2. Un modelo sin paradojas

Alfred R. Mele ha propuesto un modelo alternativo que él mismo ha descrito como deflacionista y que también es conocido en la literatura como no-intencionalista (Fernández, “Self-deception and self-knowledge”). Neil Levy resume muy bien la ventaja que presenta este nuevo modelo frente al anterior cuando, en su libro *Neuroethics*, afirma:

[Los deflacionistas] intentan explicar el autoengaño sin postular ninguna maquinaria mental extravagante requerida por la concepción tradicional. Se deshacen de la maquinaria extra al deshacerse de los requerimientos que la necesitan: tanto el requerimiento de la intencionalidad, como el de la contradicción de creencias (260).

Al dispensar de estos dos elementos, no surgen las paradojas que debía enfrentar el modelo canónico y que lo obligan a convertirse en algo demasiado demandante y enrevesado.

¿Cómo funciona este modelo? La idea es sencilla. Para los deflacionistas el autoengaño puede ser explicado en términos de mecanismos de adquisición de creencias parcializados motivacionalmente. Para explicar esto veamos primero cuál es la estructura básica del modelo deflacionista.

S incurre en autoengaño al adquirir la creencia de que P si:

- (1) La creencia de que P que S adquiere es falsa.
- (2) S trata los datos relevantes, o al menos aparentemente relevantes, para el valor de verdad de P de una manera motivacionalmente sesgada.
- (3) Este tratamiento sesgado es una causa no desviada de que S adquiriera la creencia de que P.
- (4) Los datos a disposición de S en ese momento proveen mayor evidencia para no-P que para P (Mele, *Self-deception Unmasked*, 50)².

La primera condición establece simplemente que la creencia que el sujeto S forma es contraria a la verdad, lo cual se ve complementado con la condición 4, que estipula que S tiene más evidencia a favor de la verdad que a favor de la creencia falsa que termina formando. Estas dos condiciones son típicas de todo caso de autoengaño; el *quid* está en la segunda condición: ¿qué es tratar los datos de una manera motivacionalmente sesgada? Es, en pocas palabras, tratar la evidencia disponible de manera incorrecta gracias a la presencia de una motivación fuerte. Claro, esto no parece ser explicación suficiente, así que ilustraremos esto nuevamente con un ejemplo literario.

En “Gimpl el ingenuo” Singer presenta a Gimpl, un hombre bueno que se casa con una mujer poco agraciada y que lo trata mal. Logra, sin embargo, construir con ella una familia y tienen hermosos hijos a los que Gimpl ama con todo su corazón. Una noche Gimpl llega temprano a su casa y ve en la oscuridad de la noche que en la cama, al lado de Elka, su esposa, yace la figura de un varón. En lugar de armar un escándalo Gimpl piensa en no despertar a sus hijos y, sin decir una palabra regresa a su panadería a pasar allí la noche. Al día siguiente increpa a su esposa quien niega vehementemente todo. Acude al rabino a buscar consejo y éste le dice que debe marcharse de la casa y no volver a cruzar el camino con Elka nunca más. Gimpl lo hace, pero con el pasar de los días empieza a extrañar a su familia, desea volver a sentir el calor de sus hijos. Es este deseo, el de estar con sus hijos, el que lo lleva a reevaluar la evidencia y verla ahora bajo otra luz:

² Hemos defendido en otro lugar que esta estructura no es sólo aplicable y útil para los casos de autoengaño. Creemos que se trata de una estructura muy versátil y aplicable a otros fenómenos que usualmente se tildan de meramente irracionales (ver. Muñoz y Patarroyo, *Culpa sin trasgresión*).

Por otra parte, puesto que ella lo negaba con tanta vehemencia, ¿quizás sólo lo imaginé? Las alucinaciones ocurren. A veces uno ve una figura, un hombrecito o algo parecido, y cuando te acercas ves que no es nada, que no había nada ahí. Y si así fue, yo estaría cometiendo una injusticia contra ella. Cuando llegué a este punto en mis pensamientos, empecé a llorar. [...] A la mañana siguiente fui a ver al rabino y le dije que me había equivocado (23).

Gimpl regresa a su casa y pide perdón a Elka. Quien conozca el cuento sabe que, a fin de cuentas, Gimpl no estaba equivocado, que su esposa sí le fue infiel, y no sólo esa vez, sino otras más, y en cada una de ellas Gimpl, por el amor a sus hijos, reevaluaba la evidencia y vuelve a su casa a pedir perdón.

Nótese que aquí no hay creencias opuestas y contradictorias, pues Gimpl jamás cree simultáneamente que su esposa le es infiel y que le es fiel. De igual manera, Gimpl no tiene la intención ni el deseo de engañarse. Su manejo motivacionalmente sesgado de la evidencia es producido por su deseo de ver a sus hijos, no por el deseo de creer en la fidelidad de Elka. No hay en ningún momento un deseo ni una intención de Gimpl de urdir un plan para engañarse a sí mismo. El modelo, al prescindir de estos dos elementos que sí tenía el canónico, no ha de enfrentar entonces las paradojas.

Una ventaja adicional reside en la cantidad de evidencia empírica que hay para explicar casos de manejo motivacionalmente sesgado de evidencia³. Este manejo se puede dar de cuando menos cuatro maneras: (i) Malinterpretación negativa –cuando el deseo lleva a no contar en contra de *P* evidencia que va en su contra–, (ii) malinterpretación positiva –cuando el deseo nos lleva a interpretar evidencia que en realidad va en contra de *P* como si fuera a su favor–, (iii) enfoque selectivo –cuando el deseo nos lleva a enfocarnos sólo en la evidencia que cuenta a favor de *P* y no en aquella que va en su contra– y (iv) recolección selectiva de la evidencia –cuando el deseo nos lleva a no prestar atención a evidencia en contra de *P* fácilmente obtenible, y a recaudar evidencia a su favor más difícil de obtener– (Mele, “Have I Unmasked Self-Deception”, 261-2).

³ Un recorrido histórico que presenta una gran cantidad de experimentos hechos por una gran cantidad de autores puede encontrarse en Hann *et. al.* 2014.

Alguien podría cuestionar si no se incurre aquí en una falacia de cambio de tema, en donde al eliminar las creencias contradictorias y la intención de engañar se ha eliminado el autoengaño y se está explicando una cosa diferente. La respuesta a este cuestionamiento es sencilla pero no contundente: el modelo deflacionista pretende explicar exactamente los mismos casos que el modelo canónico; es el modelo canónico el que postula la existencia de una intención de engañarse y de creencias contradictorias, y por ello es éste el que debe demostrar la existencia de tales elementos (tarea que no ha dado frutos aún). En su lugar, el modelo deflacionista no incurre en estas postulaciones y, sin embargo, explica los casos acudiendo a una estructura para la que ya hay bastante evidencia. En palabras de Mele:

No he intentado probar que no existe el autoengaño *estricto* [refiriéndose al que hemos llamado canónico]. En lugar de ello he ofrecido un marco explicativo para dar cuenta del autoengaño como es comúnmente conocido, y he defendido que, mientras que hay una cantidad considerable de evidencia a favor de la existencia de los procesos y fenómenos a los que he apelado al construir este marco, no hay nada que siquiera se acerque a ser similar en lo referente a evidencia a favor de las ‘creencias duales’ o la ocurrencia de intentos inconscientes del tipo que mencionan los defensores del modelo tradicional (Mele, *Self-Deception Unmasked*, 124).

Así, no se trata de refutar concluyentemente un modelo y *de demostrar* la veracidad del otro. En su lugar, se trata de argumentar acerca de las dificultades que ha de enfrentar uno, y la conveniencia, simplicidad y apoyo empírico que tiene el otro y que lo hacen, en suma, más plausible.

3. El problema de la evidencia que se resiste y los vínculos sociales

Creemos que el modelo deflacionista adolece de un problema, y es que está construido sobre casos en los que la evidencia es meramente perceptual. Mele, por ejemplo, habla de cómo alguien puede leer los comentarios de un evaluador que ha rechazado un artículo académico para su publicación, y no ver en ellos la contundencia o verdad de las afirmaciones, y más bien ver injusticias y juicios apresurados (Mele, “Have I Unsmaked Self-Deception”, 261). El problema reside en que esto es muy similar a ver el color

de un objeto. Si el observador sufre de protanopia verá una silla roja como amarilla, y eso será todo porque la silla no *se resiste* a ser vista como amarilla. Pero no toda la evidencia con la que ha de lidiar una persona en condición de autoengaño es evidencia de este tipo. Queremos defender que se hace necesario dividir la evidencia en dos tipos diferentes, aquella que llamaremos *estática*, como la silla, las palabras en la carta de rechazo del artículo, etc., y la que llamaremos *dinámica*. La evidencia dinámica corresponde a personas con las que el agente se relaciona e interactúa. Esta evidencia, a diferencia de la *estática*, suele resistirse a ser juzgada de manera incorrecta. Retomemos el ejemplo del daltonico y la silla. La silla es netamente pasiva en la percepción; el agente la ve amarilla en vez de roja y eso es todo. Pero si el agente juzga que una persona es egoísta cuando ésta no lo es, esta segunda persona se resistirá, lo increpará y dirá que el agente ha de corregir su juicio inicial. Esto obliga a hacer más complejo el modelo deflacionista porque hay que explicar cómo alguien puede permanecer en condición de autoengaño pese a tener acceso y relación con evidencia *dinámica*. La resistencia que ofrece esta evidencia debería servir para *sacar* al agente de su condición de autoengaño, para corregir su malinterpretación inicial y formar entonces la creencia correcta. Sin embargo, los casos comúnmente conocidos como ejemplos de autoengaño suelen estar en presencia de evidencia *dinámica* y, sin embargo, ser resilientes, es decir, no ser corregidos por la presencia de esta evidencia. ¿Cómo es esto posible?

Deseamos defender una explicación de esto que tiene que ver con dos elementos. El primero de ellos es el costo que la corrección traería para el agente. El elemento motivacional ha de ser un elemento fuerte, arraigado, importante para el agente como para causar el tratamiento sesgado de la evidencia. Ya que se trata de algo importante para el agente, es de suponer que abandonarlo o rechazarlo trae un costo. Entre más importante sea el elemento motivacional para el agente, más posibilidades hay de hacer un manejo sesgado de la evidencia pero, a la vez, más difícil será corregir las creencias falsas que se han formado gracias a este tratamiento. Retomando el caso de Gimpl, el elemento motivacional era el deseo profundo de estar con sus hijos. Corregir la creencia falsa de que su esposa le es fiel implica que él debe (tal como le indicó el rabino) alejarse de su casa y, por

tanto, de sus hijos. Corregir la creencia implica entonces negar o no satisfacer ese deseo tan fuerte que tiene, y eso es un precio elevado a pagar.

Esto ayuda a explicar las diferencias en la resiliencia de casos de autoengaño diferentes. Puede haber casos de autoengaño en donde el elemento motivacional no sea muy fuerte, pero sólo si la evidencia es *estática*. Cuando hay también presencia de evidencia *dinámica* ha de haber una motivación fuerte, que sea costoso abandonar, para que la resistencia de esta evidencia no produzca la corrección de la creencia. Esto, sin embargo, no explica aún cómo es que se maneja esta evidencia para mantener la creencia falsa. En el caso de la evidencia estática cualquiera de las cuatro formas de manejo sesgado de la evidencia mencionadas en la sección anterior basta para formar la creencia falsa y, como la evidencia no se resiste, la creencia no se ve controvertida ni puesta a prueba. Pero la evidencia *dinámica* se resiste, y el manejo de ella no puede ser el de ninguna de las cuatro formas antes vistas porque si, por ejemplo, el sujeto incurre en malinterpretación positiva, la evidencia misma (la persona con quien está interactuando) se resistirá y le dirá que está incurriendo en un error. No basta entonces con la malinterpretación negativa (ni ninguna de las otras tres formas de manejo sesgado de la evidencia), ha de haber algo más.

Este elemento adicional es, a nuestro juicio, el aislamiento. El corte de los vínculos sociales y de las interacciones con el sujeto que se resiste, con esa persona que es ahora una evidencia dinámica. El agente que está motivacionalmente sesgado recurre entonces a un juicio rápido en el cual juzga a la otra persona como indigna, incompetente o en general no apta para tratar el tema en cuestión con ella y, al hacerlo, corta cierta relación, cierta interacción, cierto vínculo con ella. Esto es algo perfectamente común. En el día a día las discusiones políticas, por ejemplo, sirven de ejemplo de abundantes casos de reacciones de este estilo. Personas con serios sesgos acerca de sus creencias políticas terminan por decir que “con X no se puede hablar de política”, no porque realmente no se pueda hablar con él o ella, sino porque se ha resistido a acomodarse a los deseos y motivaciones que nutren las creencias erradas del primer sujeto. Decir esto viene acompañado de un corte de relaciones sociales. Ahora el agente y la persona X han perdido un vínculo. Ya no hablan más de política

porque el agente no vuelve a tratar el tema con X. Pensemos en cuántas situaciones esto se hace común, por ejemplo, los cortes de relaciones por discrepancias en opiniones acerca de arte o deportes.

Esto, por supuesto, tiene grados. El corte del vínculo puede darse sólo para discusiones en un tema específico, y en todo lo demás seguir gozando de buena comunicación; o bien puede ser severo e implicar un corte en las relaciones con el otro en general.

Un ejemplo claro de este corte de relaciones lo representa el personaje de Timón de Atenas, de la obra de Shakespeare que lleva el mismo nombre. Exuberante y sensible a la adulación, Timón ofrece banquetes costosos y da regalos y dinero a sus amigos. Tanta generosidad termina por vaciar sus arcas; y los intentos de Flavio, su mayordomo y administrador de sus bienes, por mostrarle la verdadera situación son en vano. Es el filósofo Apemanto quien lo enfrenta e intenta mostrarle que tantos elogios y vanaglorias de sus supuestos amigos lo han enceguecido. Y es justamente allí que Timón corta la relación con él:

Timón: Apemanto, si tú no fueras hosco, yo querría portarme bien contigo.

Apemanto: No, no quiero nada; y si yo cayera también bajo el soborno, no quedaría nadie que despotricara contra ti, y entonces pecarías más de prisa. Tanto tiempo llevas dando, Timón, que temo que pronto te darás tú mismo en papel.

Timón: No, una vez que empezáis a despotricar contra la sociedad, he jurado no haceros caso. Adiós, y volved con mejor música.

Apemanto: Muy bien, si ahora no me oyes, después ya no podrás. [...] Ah, ¿por qué los oídos de los humanos son sordos a los consejos y no a la adulación? (Acto I, Escena II, 93).

Nuestra posición se resume de esta manera: ante la evidencia estática basta la malinterpretación; ante la evidencia dinámica se recurre además al sacrificio de vínculos sociales.

Esto permite explicar el comportamiento social de las personas autoengañadas y, lejos de ser algo debilitante para el modelo deflacionista, creemos que centrar la mirada sobre la evidencia dinámica termina por enriquecerlo y abre la puerta para otras investigaciones de las posibles consecuencias que

pueda traer esto que hemos propuesto. No tenemos el espacio para desarrollar un análisis de ellas, pero bien vale la pena enunciar una como punto digno de un futuro desarrollo.

En casos de autoengaño severo, es decir, donde la motivación es muy fuerte y, por ende, es costoso corregir la creencia errada, es concebible ver al agente como un inadecuado juez de la verdad. Esto es, no es claro que la salida a su situación de autoengaño vaya a darse por el examen de los *hechos* o de la evidencia *estática*. Es muy probable que la salida esté en el valor que les da a las relaciones sociales. Intentar mantener su creencia ha de llevarlo, según hemos defendido, a ir cortando lazos sociales con aquellos que forman la evidencia *dinámica*. Es perfectamente razonable asumir que las personas más cercanas y más valiosas en su vida serán las que se resistan y que sea con ellas con quienes se corten los lazos. Y es también razonable pensar que cortar estos lazos tiene un costo. La salida al autoengaño puede darse, entonces, cuando el costo de cortar estos lazos y de aislarse llega a superar el costo de corregir la creencia falsa y no satisfacer el deseo o motivación que a ella ha llevado. En pocas palabras, sería la necesidad de las relaciones sociales y su valor, lo que permitiría a un agente salir del autoengaño, y no la verdad o los hechos como algo estático.

Bibliografía

Audi, Robert. "Self-Deception, Rationalization and Reasons for Acting". *Perspectives on Self-Deception*. Ed. Okseberg-Rorty, Amélie; McLaughlin, Los Angeles, University of California Press, 1988, pp. 92-120.

Bermúdez, José Luis. "Self-Deception, Intentions and Contradictory Beliefs". *Analysis* 60.4, 2000, pp. 309-19.

BonJour, L. *Epistemology: Classic Problems and Contemporary Responses*. Oxford, Rowman & Littlefield Publishers, 2010.

Buckwalter, Wesley, David Rose, and John Turri. "Belief through Thick and Thin". *Noûs*, 49.4, 2013, pp. 748-75.

Byrne, A. *Transparency and Self-Knowledge*. Oxford, Oxford University Press, 2018.

Davidson, Donald. *Problems of Rationality*. Oxford, Oxford University Press, 2004.

Fernández, Jordi. "Self-Deception and Self-Knowledge". *Philosophical Studies*, 162.2, 2013, pp. 379-400.

Funkhouser, Eric, and David Barrett. "Robust, Unconscious Self-Deception: Strategic and Flexible". *Philosophical Psychology*, 29.5, 2016, pp. 1-15.

Greenwald, Anthony. "Self-Knowledge and Self-Deception: Further Consideration". *The Mythomanias: The Nature of Deception and Self-Deception*. Ed. Myslobodsky, Michael, New Jersey, Lawrence Erlbaum Associates Publishers, 1977, pp. 51-71.

Hahn, Ulrike, and Adam J. L. Harris. "What Does It Mean to Be Biased: Motivated Reasoning and Rationality". *Psychology of Learning and Motivation*. Ed. Ross, Brian, San Diego, Academic Press, 2014, Pp. 41-102.

Levy, N. *Neuroethics: Challenges for the 21st Century*. Cambridge, Cambridge University Press, 2007.

Mele, Alfred R. "Have I Unmasked Self-Deception or Am I Self-Deceived?". *The Philosophy of Deception*. Ed. Martin, Clancy W., Oxford, Oxford University Press, 2009, pp. 260-276.

Mele, A. R. *Self-Deception Unmasked*. Princeton, Princeton University Press, 2001.

Muñoz, Carlos A.; Patarroyo G., Carlos G. *Culpa Sin Trásgresión: Un análisis filosófico de la culpa del sobreviviente*. Bogotá, Universidad del Rosario, En prensa.

Okseberg-Rorty, Amélie. "The Deceptive Self: Liars, Layers and Lairs". *Perspectives on Self-Deception*. Ed. Okseberg-Rorty,

Amélie; McLaughlin, Brian, Los Angeles, University of California Press, 1988. Pp. 11-28.

Shakespeare, William. *Timón de Atenas*. Buenos Aires, Losada, 2004.

Singer, Isaac Bashevis. *Cuentos*. Barcelona, Lumen, 2018.

Stokke, A. *Lying and Insincerity*. Oxford, Oxford university Press, 2018.

Strandberg, H. *Self-Knowledge and Self-Deception*. Palgrave Macmillan, 2015.

Zweig, Stefan. *Novela De Ajedrez*. Barcelona, Acantilado, 2000.

Recibido: 24 de enero de 2019. Revisado: 29 de enero de 2019. Publicado: 31 de enero de 2019. *Revista Letral*, n.º 21, 2019, pp. 241-257. ISSN 1989-3302.

DOI: <http://dx.doi.org/10.30827/RL.voi21.8608>